β-Amyrin synthase from *Euphorbia tirucalli* L. Functional analyses of the highly conserved aromatic residues Phe413, Tyr259 and Trp257 disclose the importance of the appropriate steric bulk, and cation-π and CH-π interactions for the efficient catalysis of the polyolefin cyclization cascade

Ryousuke Ito, Chika Nakada and Tsutomu Hoshino*

Graduate School of Science and Technology and Department of Applied Biological Chemistry, Faculty of Agriculture, Niigata University, Ikarashi 2-8050, Nishi-ku, Niigata 950-2181, Japan
Figure S1. Amino acid alignment of various oxidosqualene cyclases
D485C486TAE involved in EtAS β-amyrin synthase

**EtAS F728**


**SHC**

The triterpene cyclases amino acids sequences were aligned using Clustal W, as implemented in the (CLC Sequence Viewer (CLC bio), and the figure was made by GenDoc (http://www.nrbstc.org/gfx/genedoc/). EtAS: *Euphorbia tirucalli* β-amyrin synthase (AB206469), PNY1: *Panax ginseng* β-amyrin synthase (AB009030), BPY: *Betula platyphylla* β-amyrin synthase (AB055512), PSY: *Pisum sativum* β-amyrin synthase (AB034802), AtLUP1: *Arabidopsis thaliana* multifunctional triterpene cyclase (At1g78970), TRW: *Taraxacum officinale* lupeol synthase (AB025345), OEW: *Olea europaea* lupeol synthase (AB025343), BPW: *Betula platyphylla* lupeol synthase (AB055511), AtCAS1: *Arabidopsis thaliana* cycloartenol synthase (At2g07050), PsCAS: *Pisum sativum* cycloartenol synthase (D89619), PgCAS: *Panax ginseng* cycloartenol synthase (AB009029), BpCAS1: *Betula platyphylla* cycloartenol synthase (AB055509), AtLAS1: *Arabidopsis thaliana* lanosterol synthase (At3g45130), HsLAS: *Homo sapiens* lanosterol synthase (P48449), MmLAS: *Mus musculus* lanosterol synthase (AK044016), ScLAS: *Saccharomyces cerevisiae* lanosterol synthase (P38604), SHC: *Alicyclobacillus acidocaldarius* squalene-hopene cyclase.
Fig. S2. EIMS and NMR spectra of Product 10-acetate

Fig. S2-1. EIMS spectrum of Product 10 acetate

Fig. S2-2. $^1$H-NMR spectrum of product 10-Ac in CDCl$_3$ (600 MHz).
Fig. S2-3. $^{13}$C-NMR spectrum of product 10 Ac (150 MHz)

Fig. S2-4. $^1$H-$^1$H COSY 90 of product 10 Ac
Fig. S2-5. HOHAHA spectrum of product 10 Ac

Fig. S2-6. NOESY spectrum of product 10 Ac
Fig. S2-7. HSQC spectrum of product 10 Ac

Fig. S2-8. HMBC spectrum of product 10 Ac
The stereoisomer of 10, described below, was isolated by Wu’s group from the mutants of S. cerevisiae lanosterol synthase, which was reported in the following references.


Fig. S3. EIMS and NMR spectra of Product 12-acetate.

Fig. S3-1. EIMS spectrum of product 12-Ac

Fig. S3-2. $^1$H-NMR spectrum of product 12-Ac in C$_6$D$_6$ (400 MHz).
Fig. S3-3. $^{13}$C-NMR spectrum of product 12-Ac in C$_6$D$_6$ (100 MHz).

Fig. S3-4. $^1$H-$^1$H COSY spectrum of product 12-Ac in C$_6$D$_6$. 
Fig. S3-5. HOHAHA spectrum of product 12-Ac in C₆D₆.

Fig. S3-6. NOESY spectrum of product 12-Ac in C₆D₆ (400 MHz).
Fig. S3-7. NOESY spectrum of product 12-Ac in C₆D₆ (600 MHz)-Expanded Region

Fig. S3-8. HSQC spectrum of product 12-Ac in C₆D₆
Fig. S3-9. HMBC spectrum of product 12-Ac in C6D6

Fig. S3-10. NMR data analyses for proposing structure of 12-Ac, optical rotation and HRMS data.
Fig. S4. EIMS and NMR spectra of product 15 acetate.

Fig. S4-1. EIMS spectrum of 15-Ac.

Fig. S4-2. $^1$H-NMR spectrum of 15-Ac in CDCl$_3$ (400 MHz).
Fig. S4-3. $^{13}$C NMR spectrum of 15-Ac in CDCl$_3$ (100 MHz).

Fig. S4-4. $^1$H-$^1$H COSY spectrum of 15-Ac in CDCl$_3$. 
Fig. S4-5. HOHAHA spectrum of 15-Ac in CDCl₃.

Fig. S4-6. NOESY spectrum of 15-Ac in CDCl₃.
Fig. S4-7. HSQC spectrum of 15-Ac in CDCl₃.

Fig. S4-8. HMBC spectrum of 15-Ac in CDCl₃.
Fig. S4-9. NMR data analyses for proposing structure of 15-Ac, optical rotation and HRMS data.

Dammara-20(Z),24-dien-3β-ol

![Chemical Structure Image]

400 MHz in CDCl3
the solvent peak 1H: 7.26 ppm; 13C: 77.0 ppm

<table>
<thead>
<tr>
<th>NO.</th>
<th>1H</th>
<th>13C</th>
<th>NO.</th>
<th>1H</th>
<th>13C</th>
<th>NO.</th>
<th>1H</th>
<th>13C</th>
<th>NO.</th>
<th>1H</th>
<th>13C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.06 (m)</td>
<td>38.77</td>
<td>9</td>
<td>1.34 (m)</td>
<td>50.88</td>
<td>17</td>
<td>2.73 (1H, m)</td>
<td>41.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.64 (2H, m)</td>
<td>23.71</td>
<td>10</td>
<td>—</td>
<td>—</td>
<td>18</td>
<td>0.88 (3H, t)</td>
<td>15.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.40 (4H, d=2.3, 6.4 Hz)</td>
<td>80.93</td>
<td>11</td>
<td>1.17 (m); 1.62 (m)</td>
<td>21.57</td>
<td>19</td>
<td>0.69 (3H, t)</td>
<td>16.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>37.90</td>
<td>12</td>
<td>1.06 (m); 1.47 (m)</td>
<td>24.88</td>
<td>20</td>
<td>—</td>
<td>135.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.84 (1H, m)</td>
<td>58.96</td>
<td>13</td>
<td>1.62 (1H, m)</td>
<td>44.44</td>
<td>21</td>
<td>1.64 (3H, s)</td>
<td>19.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.55 (2H, m)</td>
<td>18.16</td>
<td>14</td>
<td>—</td>
<td>40.32</td>
<td>22</td>
<td>5.15 (2H, J=7.2 Hz)</td>
<td>124.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.29 (m); 1.50 (m)</td>
<td>36.38</td>
<td>15</td>
<td>1.12 (m); 1.50 (m)</td>
<td>31.81</td>
<td>23</td>
<td>2.66 (2H, m)</td>
<td>26.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>40.53</td>
<td>16</td>
<td>1.45 (m); 1.70 (m)</td>
<td>26.55</td>
<td>24</td>
<td>5.05 (2H, J=7.2 Hz)</td>
<td>123.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M⁺: Obsd. 468.39780; Calcd. 468.39673
[α]D20 = 463.3 (c=0.035, CHCl3)
Fig. S5. EIMS and NMR spectra (400 MHz, C₆D₆) of product 17-acetate

Fig. S5-1. EIMS spectrum of product 17-Ac.

Fig. S5-2. ¹H-NMR spectrum of product 17-Ac (400 MHz, C₆D₆).
Fig. S5-3. $^{13}$C-NMR spectrum of product 17-Ac (100 MHz, C$_6$D$_6$).

Fig. S5-4. $^1$H-$^1$H COSY spectrum of product 17-Ac.
Fig. S5-5. HOHAHA spectrum of product 17-Ac.

Fig. S5-6. NOESY spectrum of product 17-Ac.
Fig. S5-7. HSQC spectrum of product 17-Ac.

Fig. S5-8. HMBC spectrum of product 17-Ac.
**Fig. S5-9.** NMR data analyses for proposing structure of 17-Ac and HREIMS data. Reliable optical rotation was not obtained due to the sample loss during isolation and the instrumental analyses.

Product 17 acetate in C$_6$D$_6$

400 MHz in C$_6$D$_6$

<table>
<thead>
<tr>
<th>NO.</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>NO.</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>NO.</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
<th>NO.</th>
<th>$^1$H</th>
<th>$^{13}$C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.62 (m); 1.52 (m)</td>
<td>38.56</td>
<td>9</td>
<td>1.43 (dd, 12.9 Hz)</td>
<td>50.56</td>
<td>17</td>
<td>146.2</td>
<td>25</td>
<td>121.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.71 (q); 1.84(m)</td>
<td>24.09</td>
<td>10</td>
<td>—</td>
<td>—</td>
<td>18</td>
<td>1.038 (3H, s)</td>
<td>18.29</td>
<td>26</td>
<td>1.822 (3H, s)</td>
<td>25.85</td>
</tr>
<tr>
<td>3</td>
<td>4.80 (s); J=11.5, 4.8 Hz</td>
<td>80.44</td>
<td>11</td>
<td>1.22 (m); 1.55(m)</td>
<td>24.03</td>
<td>19</td>
<td>0.876 (3H, s)</td>
<td>16.29</td>
<td>27</td>
<td>1.713 (3H, s)</td>
<td>17.80</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>38.09</td>
<td>12</td>
<td>1.63 (m); 2.14 (p); J=12.5 Hz</td>
<td>21.96</td>
<td>20</td>
<td>145.7</td>
<td>28</td>
<td>0.19 (3H, s)</td>
<td>28.08</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.83 (m)</td>
<td>56.24</td>
<td>13</td>
<td>3.05 (1H, m); J=13.5, 3.2 Hz</td>
<td>49.37</td>
<td>21</td>
<td>5.23 (s); 5.14 (s)</td>
<td>110.5</td>
<td>29</td>
<td>1.038 (3H, s)</td>
<td>16.26</td>
</tr>
<tr>
<td>6</td>
<td>1.46 (m); 1.56(m)</td>
<td>18.47</td>
<td>14</td>
<td>—</td>
<td>53.45</td>
<td>22</td>
<td>2.47 (m); 2.37(m)</td>
<td>35.44</td>
<td>30</td>
<td>1.247 (3H, s)</td>
<td>16.12</td>
</tr>
<tr>
<td>7</td>
<td>1.91 (dd, J=12.9, 2.4 Hz); 1.58(m)</td>
<td>35.46</td>
<td>15</td>
<td>1.84 (m); 2.48 (m)</td>
<td>39.00</td>
<td>23</td>
<td>2.45 (2H, m)</td>
<td>28.02</td>
<td>31</td>
<td>—</td>
<td>170.0</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>40.10</td>
<td>16</td>
<td>5.84 (pt)</td>
<td>127.1</td>
<td>24</td>
<td>5.42 (very broad s)</td>
<td>124.9</td>
<td>32</td>
<td>1.882 (3H, s)</td>
<td>20.86</td>
</tr>
</tbody>
</table>

* and $^b$: The assignments are exchangeable between the same letters.
Fig. S6. GCMS spectrum of the acetate of the hexane-extract obtained from the Y259H mutant. 19 acetate. The EIMS of 19 acetate was almost identical to that of dammara-(E)-20(22),(E)24(25)dien-3β-ol acetate that was isolated from F728H mutant (Ito, R.; Hashimoto I.; Masukawa Y.; Hoshino, T. Chem.-Eur. J. 2013, 19, 17150-17158, see Supporting Information, Fig. S36, EIMS) and is quite similar as that of EIMS of 15-acetate.

[Diagram of lanosterol molecule with green color and annotations]

Lanosterol molecule is shown with green color. We have reported the detailed functional analysis of F728 in the previous paper (R. Ito, I. Hashimoto, Y. Masukawa and T. Hoshino, Chem. Eur. J. 2013, 19, 17150-17158). The site-directed mutagenesis experiments have demonstrated that the function of cation/π interaction was assigned to the F728 residue. On the other hand, the steric bulk at 474 position is critical to the accurate folding of oxidosqualene to complete the polycyclization reaction (Ito, R.; Masukawa, Y.; Nakada, C.; Amari, K.; Nakano, C.; Hoshino, T. Org. Biomol. Chem. 2014, 12, 3836-3846). By the mutagenesis experiments, we found that D485C486TA motif triggers the polycyclization reaction and the C564 is involved in hydrogen bond formation with the carboxyl residue of D485, resulting in enhancement of the acidity (R. Ito, Y. Masukawa and T. Hoshino, FEBS J., 2013;280:1267-1280). F474 is situated in the vicinity to B-ring formation site. F413 residue is located in approximate to the C/D ring. Y259 and W257 that correspond to Y261and to W259, respectively, of P. ginseng PNY β-amyrin synthase, are also marked in this model. It was reported that the Y261H mutant of PNY gave the teracyclic products (see the Text and the ref. T. Kushiro, M. Shibuya, K. Masuda, Y. Ebizuka, J. Am. Chem. Soc., 2000, 122, 6816-6824). The hydrogen bonding may occur between OH of the polar amino acids (Ser, Thr or Tyr mutants) and the phenolic OH of Y259, because of the proximal distance between them (see Fig. S7.2), which would have brought about the inappropriate placement of Y259, resulting in the decreased activities of the polar amino acid-substituted mutants (see Figure 3 in Text).
**Fig. S7.2.** Distances between OH of Y259 and OH of polar amino acids (Ser and Thr), OH of Tyr and NH or N of His, which were estimated from the modeling constructed by the methods of ESyPred3D. Lanosterol molecule is shown with green color.

Hydrogen bonding is presumed between OH of Y259 and the polar group of Ser, Thr, His and Tyr, and water molecule(s) is also likely to intervene between the OH of Y259 and the polar groups of the substituted amino acids.
**Fig. S8.** GC traces of the lipophilic acetylated materials produced by various mutants targeted for F413. A 100 mL culture of each mutants were subjected to centrifugation. The cell pellets were subjected to saponification with 15% KOH/MeOH under reflux condition, followed by the extraction of the lipophilic materials with hexane extract (3 x 10 mL). The triterpene fraction including products was obtained by partial purification with a SiO$_2$ column to remove oxidosqualene, dioxidosqualene and nontriterpene impurities (hexane/EtOAc=100:1), followed by acetylation with Ac$_2$O/Py. The acetate mixture was dissolved in 1.0 mL of hexane. A 0.5 μL of the hexane was injected to the GC apparatus. The GC conditions were as follows: J & W, DB-1 capillary column (Length 30 m, I.D. 0.32 mm, Film Thickness 0.25 mm); column injection temp., 300°C; column temperature, 245-270°C (0.35 °C/min).
Fig. S9. Estimation of the EtAS enzymatic activities for the wild-type and the F413X mutants.

<Expression level of β–amyrin synthase (EtAS)>

![Expression level (mg/L culture) of the EtAS enzymes of the wild-type and the site-directed mutant. The yeast cells grown in 1L-medium were collected and the protein amounts (mg) were quantified by Western blot analysis.]

![GC analyses for the quantities of oleanane-type triterpenes>]

Fig. S9.2. The quantities of products 2 and 12 (oleanane skeleton) produced by 1L-culture of each of the mutant strains that were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard.
<Enzyme activities for the production of oleanane-type triterpenes>

**Fig. S9.3.** Enzyme activities of the mutants relative to that of the wild-type for the production of oleanane skeleton. The wild-type activity (100%) indicates the sum of the relative activities shown in Figs. S9.3, Fig. S9.5 and S9.7. These enzyme activities were estimated by dividing the amounts of oleanane-type products (2 and 12) by the expressed quantities of EtAS enzymes. This means that the values of Fig. S9.2 were divided by those of Fig. S9.1. The wild-type did not show 100% activity, i.e., 96.3±0.7 %, because the wild-type produced the tetracycles 13 and 14 in a small amount (see Figs. S9.5 and S9.7).
< GC analyses for the quantities of Dammarane-types>

**Fig. S9.4.** The quantities of products 10, 11, 13 (dammarane skeleton) produced by 1L-culture of each of the mutant strains that were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard. Product 10 is tentatively categorized as Dammarane skeleton, but not 17-epi-dammarane skeleton. To make sense, the stereochemistry of C-20 must be determined.

< Enzyme activities for the production of Dammarane-types>

**Figure S9.5.** Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of dammarane-type products (10, 11, 13) by the expressed quantities of EtAS enzymes, that is, the values of Fig. S9.4 were divided by those of Fig. S9.1. The activity of the wild-type for the production of 13 and 14 was very low, i.e., 1.53±0.61%, thus, the relative activity does not correspond to 100%. The total values of Fig.S9.3 and Fig. S9.5 for the wild-type corresponds to 100%.
< GC analyses for the quantities of 17-epi-Dammarane-type>

Fig. S9.6. The quantities of products 14 (17-epi-dammarane skeleton) produced by 1L-culture of each of the mutant strains, which were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard.

< Enzyme activities for the production of 17-epi-Dammarane-types>

Figure S9.7. Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of 17-epi-dammarane-type product 14 by the expressed quantities of EtAS enzymes (%).
< GC analyses for the quantities of bicyclic product >

**Fig. S9.8.** The quantities of products 9 produced by 1L-culture of each of the mutant strains, which were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard.

**Figure S9.9.** Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of bicyclic product 9 by the expressed quantities of EtAS enzymes (%).
Fig. S10. Estimation of the EtAS enzyme activities for the Y259X mutants.

<Expression level of β-amyrin synthase (EtAS)>  

![Expression level of EtAS enzymes](image)

**Fig. S10.1.** Expression level (mg/L) of the EtAS enzymes of the wild-type and the site-directed mutant. The yeast cells grown in 1L-medium were collected and the protein amounts (mg) were quantified by Western blot analysis.

<GC analyses for the quantities of oleanane-type triterpenes>

![GC analyses for triterpenes](image)

**Fig. S10.2.** The quantities of products 2, 16 and 18 (oleanane skeleton) produced by 1L-culture of each of the mutant strains that were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard.
<Enzyme activities for the production of oleanane-type triterpenes>

Fig. S10.3. Enzyme activities of the mutants relative to that of the wild-type for the production of oleanane skeleton. See the legend to Fig. S9.3 for the calculation method.

<GC analyses for the quantities of lupanyl-type triterpene, lupeol>

Fig. S10.4. The quantities of product 20 (lupeol, 6/6/6/6/5-fused pentacycle) produced by 1L-culture of each of the mutant strains that were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard.
**<Enzyme activities for the production of lupeol>**

**Fig. S10.5.** Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of lupanyl-type product 20 by the expressed quantities of EtAS enzymes.

**<GC analyses for the quantities of tetracyclic triterpenes>**

**Fig. S10.6.** The total quantities of products 11, 13, 14, 15, 17 and 19 produced by 1L-culture of each of the mutant strains that were determined by GC analyses using GGOH (geranylgeraniol) as an internal standard.
**Enzyme activities for the production of dammarenyl skeleton**

![Graph showing enzyme activities of mutants relative to wild-type.](image)

**Fig. S10.7.** Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of tetracyclic products 11, 13, 14, 15, 17 and 19 by the expressed quantities of EtAS enzymes.

**Fig. S11.** Estimation of the EtAS enzyme activities for the W257X mutants.

**<Expression level of β-amyrin synthase (EtAS)>**

![Graph showing expression level of EtAS for wild-type and mutants.](image)

**Fig. S11.1.** Expression level of β-amyrin synthase (EtAS) for the wild-type and the mutants. The protein amounts (mg) were quantified by Western blot analysis.
**<GC analyses for the quantities of oleanane-type triterpenes>**

**Oleananyl skeleton**

![Bar graph showing product production for wild-type and variants](chart1)

**Fig. S11.2.** The quantities of oleanyl products (2 and 16) for wild-type and the variants, which were determined by GC analyses.

**<Enzyme activities for the production of oleanane-type triterpenes>**

**Oleanane**

![Bar graph showing enzyme activities for mutants relative to wild-type](chart2)

**Fig. S11.3.** Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of oleany products 2 and 16 by the expressed quantities of EtAS enzymes.
<GC analyses for the quantities of lupanyl-type triterpene, lupeol>

Fig. S11.4. The quantities of lupane skeleton (20, lupeol) for wild-type and the variants, which were determined by GC analyses.

<Enzyme activities for the production of lupeol>

Fig. 11.5. Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of lupeol 20 by the expressed quantities of EtAS enzymes.
<GC analyses for the quantities of tetracyclic triterpenes>

![Graph showing GC analyses for the quantities of tetracyclic triterpenes]

**Fig. S11.6.** The quantities of dammarenyl skeleton (13 and 14) for wild-type and the variants, which were determined by GC analyses.

<Enzyme activities for the production of dammarenyl skeleton>

![Graph showing enzyme activities for the production of dammarenyl skeleton]

**Fig. S11.7.** Enzyme activities of the mutants relative to that of the wild-type. These enzyme activities were estimated by dividing the amounts of dammarenyl skeleton (13 and 14) by the expressed quantities of EtAS enzymes.